Computer Aided Construction of Fractional Replicates from Large Factorials

by

Walter T. Federer
Charles E. McCulloch

and

Steve C. Wang

Biometrics Unit and Statistics Center
Cornell University
Ithaca, NY 14853

BU-1151-M

January 1992

Abstract

During the course of a recent statistical consultation, an investigator wished to obtain fractional factorial designs of high resolution for a large number of factors. Further, for ease of analysis, orthogonal plans were desired. Plans for such designs were not available in the literature and when the number of generators for a defining contrast exceeds two or three, the process of constructing a design is tedious. We demonstrate the use of a computer program for constructing fractions of 2^{14} and 3^{11} designs.

1. Introduction

Constructing resolution r fractional replicates from p^n factorials involves several tedious and time consuming steps, especially for large n. Appropriate use of present-day computer software and hardware can do much to alleviate the tedium of constructing such fractional replicates. We shall illustrate this with examples for p a prime power, shall discuss an inefficient method of constructing fractions for any factorial, and shall present computer software useful in this context.

If we denote a full saturated factorial of cardinality N in matrix notation as

$$\mathbf{X}_{N\times N}\beta_{N\times 1} = \mathrm{E}[\mathbf{Y}_{N\times 1}], \tag{1}$$

we may partition $\mathbf{Y}_{N\times 1}$ into any two subsets $\mathbf{Y}_1$ and $\mathbf{Y}_2$. There are corresponding partitions of $\mathbf{X}$ and β as follows:

$$\mathbf{X}\beta = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{X}_{21} & \mathbf{X}_{22} \end{bmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \mathrm{E}\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} = \mathrm{E}[\mathbf{Y}]\,. \tag{2}$$

Then using the fraction corresponding to $\mathbf{Y}_1$,

$$\mathbf{X}_{11}\beta_1 + \mathbf{X}_{12}\beta_2 = \mathrm{E}[\mathbf{Y}_1] \tag{3}$$

or

$$\beta_1 + (\mathbf{X}_{11}'\mathbf{X}_{11})^{-1}\mathbf{X}_{11}'\mathbf{X}_{12}\beta_2 = (\mathbf{X}_{11}'\mathbf{X}_{11})^{-1}\mathbf{X}_{11}'\mathrm{E}[\mathbf{Y}_1]\,. \tag{4}$$

If $(\mathbf{X}_{11}'\mathbf{X}_{11})$ does not have an inverse, then not all parameters in β_1 are estimable. β_2 is usually constructed to contain all parameters presumed to be zero or at least negligible (see, e.g., Raktoe, Hedayat and Federer, 1981). The matrix $(\mathbf{X}_{11}'\mathbf{X}_{11})^{-1}\mathbf{X}_{11}'\mathbf{X}_{12}$, or $\mathbf{X}_{11}^{-1}\mathbf{X}_{12}$ when $\mathbf{X}_{11}$ is square, is called the aliasing matrix and defines which effects in β_2 are partially or completely confounded with those in β_1. For orthogonal fractions, effects in β_2 will be completely and uniquely confounded with effects in β_1. For nonorthogonal fractions, there will be partial and some complete confounding of effects. The mean will usually be the first effect on β_1. The mean and all effects confounded with it are denoted as the defining contrast. For orthogonal fractions from prime-powered factorials, the defining contrast may be used to construct the aliasing matrix. A set of k generators is used to construct a $1/p^k$ fraction of a p^n factorial. These generators and their generalized interaction terms, in addition to the mean effect, make up the defining contrast. It should be noted that if two or more generators are completely confounded with the mean, then so are their interactions.

If the elements of β_1 contain the mean and all main effects, it is denoted as a resolution III fractional replicate or a main-effect plan. If the parameters in β_1 are the mean, all main effects, and sums of two or more two-factor interaction effects, the design is called a resolution IV plan and so forth. For a formal definition see Raktoe, Hedayat and Federer (1981, p. 88).

In a p^n factorial, the number of single-degree-of-freedom parameters in β from (1) is:

1 mean

$n(p-1)$ main effects,

$(p-1)^2 n(n-1)/2$ two-factor interaction effects,

$(p-1)^3 n(n-1)(n-2)/2(3)$ three-factor interaction effects,

$(p-1)^4 n(n-1)(n-2)(n-3)/2(3)(4)$ four-factor interaction effects,

$\vdots$

$(p-1)^n$ n-factor interaction effects.

2. Steps Involved in Constructing Fractional Replicates

As may be noted from equation (2) and the preceding section, there are numerous types of fractional replicates. In the next three sections, we shall consider only methods for orthogonal fractions from prime-power factorial designs (see Kempthorne 1947, 1952; Federer 1955). Our examples use prime numbered factorials but the method can be extended to prime-power factorials. The various steps will be illustrated with a small example, i.e., $p^n = 2^5$.

Step 1. Determine fraction $1/p^k$ and resolution r desired. In some cases, an investigator may wish to estimate main effects only, and hence a resolution III plan would be used.

Step 2. Determine the generators to be in the defining contrast. In selecting generators, the number of factors in any generator must be large enough to assure the desired resolution. For example, for a resolution V plan, no generator with three or four factors can be used, but the five-factor interaction effect, ABCDE say, would suffice to give the following 1/2 fraction of resolution V:

M = ABCDE	AD = BCE
A = BCDE	AE = BCD
B = ACDE	BC = ADE
C = ABDE	BD = ACE
D = ABCE	BE = ACD
E = ABCD	CD = ABE
AB = CDE	CE = ABD
AC = BDE	DE = ABC

The equal sign above means completely confounded with. Either the combinations in $(ABCDE)_0$ or in $(ABCDE)_1$ (see Kempthorne, 1952 or Federer, 1955 for use of this notation) would result in the 16 combinations for the 1/2 fraction.

Step 3. Determine the interaction effects among the generators and determine if any of the interaction effects have fewer than the number of desired factors. For one, two, or even three generators, this poses little problem for p = 2 or 3. However, for more generators, finding all interaction terms in the defining contrast can be a tedious process and the chances of making errors increases with number of effects. A computer program, Wylie, has been written (see Section 6) to find the interactions of the generators and to check on the number of factors in an interaction. To illustrate, consider that we wish a $1/2^2$ fraction of a 2^5. If we select the generators ABC and CDE, the generalized interaction will be ABDE which has four factors.

Step 4. Determine the aliasing structure given the defining contrast. For the example in Step 3 using the defining contrast M = ABC = CDE = ABDE, the aliasing structure would be:

$$
\begin{aligned}
M &= ABC &&= CDE &&= ABDE\\
A &= BC &&= ACDE &&= BDE\\
B &= AC &&= BCDE &&= ADE\\
C &= AB &&= DE &&= ABCDE\\
D &= ABCD &&= CE &&= ABE\\
E &= ABCE &&= CD &&= ABD
\end{aligned}
$$

There are two more lines needed to complete the aliasing structure. Any effect from the 2^5 not appearing in the first six lines is a candidate. Note that AD and AE do not appear. Hence,

$$
\begin{aligned}
AD &= BCD = ACE = BE\\
AE &= BCE = ACD = BD
\end{aligned}
$$

Note that this is a resolution III plan and part of a resolution V plan (last two lines). The computer program described in Section 6 will calculate the aliasing structure. In addition, the program allows the inclusion of only those interaction effects with f, say, or fewer factors. Thus in the above, one could include only aliases with two factors and omit all those with three or more factors from the aliasing scheme.

Step 5. Check the obtained aliasing structure to ascertain that the desired resolution is attained. Here again the computer program can be used to determine whether or not the resolution desired has been obtained.

Step 6. Construct the fractional replicate. Various combinations of levels can be used to obtain a fractional replicate. For example, for a 1/4 fraction of a 2^5, the fractions are:

$$(ABC)_0(CDE)_0 \qquad (ABC)_1(CDE)_0$$

$$(ABC)_0(CDE)_1 \qquad (ABC)_1(CDE)_1$$

Any of the above four result in the same β_1 and same aliasing structure.

Step 7. Obtain a random ordering of the combinations in the fractional replicate.

3. 2^{-k} Fractions of a 2^{14} Factorial

An investigator desired to study 14 factors each at two levels. He wished to know the minimum number of combinations required for resolution III to VII orthogonal plans. For a resolution III plan there were 1 + 14 = 15 effects to be estimated and a 2^{-10} fraction of 2^{14}, or a 2^{14-10} plan, is readily available and easily constructed. For a resolution IV plan there are 32 = 1 + 14 + 17 sums of two factor interaction effects to be estimated; this can be handled in a 2^{14-9} plan with the following nine generators: ABCF, ABDG, ABEH, ACDI, ACEJ, ADEK, BCDL, BCEM, AND BDEN (D.A. Anderson, personal communication). This design has $2^5 = 32$ runs to estimate the effects.

For a resolution V plan, there are 1 + 14 + 14(13)/2 = 106 effects to be estimated, which should be possible in $2^7 = 128$ runs, but this plan was not found. A resolution VI plan has 1 + 14 + 14(13)/2 + sums of three factor interaction effects; this would require 2^8 runs for a 2^{14-6} plan in order to estimate main effects, two factor interaction effects, and sums of three factor interaction effects. The following five generators result in a 2^{14-5} orthogonal resolution VI plan: ABCDEF, ABGHIJ, ABKLMN, CDGHKL, and CEGIKM, but a 2^{14-6} plan should be possible. Likewise, a resolution VII plan requiring 2^9 runs for the 470 effects should be possible.

The computer program in Section 6 was found to be invaluable in obtaining the above results. For the resolution IV plan above, there were 8 generators involving 247 interaction terms. The number of factors in the aliasing contrast must be at least four. The computer program lists the order of confounded factors, allowing a quick check for all effects in the aliasing contrast.

In analyzing such orthogonal fractions, it may be possible to pool the contrasts not specified by the resolution as an error term. Alternatively, the procedure known as half normal probability plots by Daniel (1959) may be used to check for large effects and to obtain an estimate of the error variance. Since these are orthogonal fractions, any estimated effect is a linear contrast of arithmetic means.

4. 3^{-k} Fractions of a 3^{11} Factorial

The investigator also desired designs involving 11 factors each at three levels. A saturated resolution VII plan would require at least $1 + 2(11) + 4(11)(10)/2 + 8(11)(10)(9)/2(3) = 1{,}563$ runs. The nearest orthogonal design would be $3^7 = 2{,}187$ runs. This number was too large for him to handle. A resolution II design would require $3^6 = 729$ runs. He decided that a resolution V design would have to suffice as it would require 243 runs since there are $1 + 2(11) + 4(11)(10)/2 = 243$ mean, main effect, and two-factor interaction single degree of freedom contrasts to be estimated. After a computer search aided by a suggestion from D.A. Anderson (pers. comm.) the following generators were found to obtain a saturated orthogonal resolution V plan: ABC^2D^2F, AB^2C^2EG, AB^2DE^2H, ACD^2E^2I, BC^2DE^2J and ABCDEK. Note that this plan is unique in that it is a saturated resolution V plan.

The program described in Section 6 was invaluable for checking that all effects in the defining contrasts contained at least five factors which is necessary for a resolution V plan. Checking a set of six generators and their interactions by hand required approximately six hours whereas it was a matter of minutes with the computer program.

For the investigation contemplated and for non-saturated designs, it may be possible to use three- and higher-factor contrasts as an error term. Alternatively a half normal probability plot procedure could be used where the square roots of the single degree of freedom sum of squares in an analysis of variance are the variates plotted. Thus large effects are located and an error term is provided.

In the actual investigation, due to the highly fractionated nature of the design, some replication was advised to obtain an estimate of pure error. This was achieved by simultaneously setting all the factors at low, at medium or at high (they were all expected to cause an effect in the same direction).

5. Saturated Resolution VII Nonorthogonal Plans

A one-at-a-time method for constructing fractional replicates is described in Anderson and Federer (1975); a saturated fraction of any resolution may easily be constructed. For a resolution VII saturated fraction from a 2^{14}, for example, there will be a mean effect, 14 main effects, 91 two-factor interaction effects, and 364 three-factor interaction effects to make a total of 470 effects. The 470 combinations may be obtained as follows:

Main effects and mean – 15 combinations:

	Factor and Level of Factor													
observation	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	1	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	1	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	1	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	1	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	1	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	1	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	1	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	1	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	1	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	1	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	1	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Two-factor interactions – 91 combinations:

16	1	1	0	0	0	0	0	0	0	0	0	0	0	0
17	1	0	1	0	0	0	0	0	0	0	0	0	0	0
18	1	0	0	1	0	0	0	0	0	0	0	0	0	0
⋮				⋮										
106	0	0	0	0	0	0	0	0	0	0	0	0	1	1

Three-factor interactions — 364 combinations:

107	1 1 1 0 0 0 0 0 0 0 0 0 0 0 0
108	1 1 0 1 0 0 0 0 0 0 0 0 0 0 0
109	1 1 0 0 1 0 0 0 0 0 0 0 0 0 0
⋮	⋮
470	0 0 0 0 0 0 0 0 0 0 0 0 1 1 1

The above method of constructing fractional replicates is the worst (in the sense of maximal variance of estimated effects) one can do for a main effect plan, i.e., the first 15 combinations (Anderson and Federer 1975). There is only once combination at level one and there are 14 at level zero. For the first 106 combinations to estimate main effects and two-factor interactions, there are 14 combinations at level one and 92 at level zero, or roughly a 1 : 6.5 ratio. For the 470 combinations each factor will appear $1 + 13 + 78 = 92$ times at the one level and 378 times at the zero level, i.e., a ratio of about 1 : 4. As the resolution of these saturated fractions approaches the complete factorial, the ratio of ones to zeros approaches a 1 : 1 ratio and the fraction approaches full efficiency.

These nonorthogonal fractions and the statistical analysis would require the inversion of a 470 by 470 matrix, a formidable task. Hence, although it was tempting to use the above fraction, the investigator opted to use an orthogonal fraction.

The one-at-a-time method of constructing a saturated main effect plan to estimate main effects and two-factor interactions from a 3^{11} factorial is as follows:

Mean and main effects — 23 combinations:

	Factor and Level of Factor										
observation	A	B	C	D	E	F	G	H	I	J	K
1	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0
3	2	0	0	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	0	0	0
5	0	2	0	0	0	0	0	0	0	0	0
6	0	0	1	0	0	0	0	0	0	0	0
7	0	0	2	0	0	0	0	0	0	0	0
8	0	0	0	1	0	0	0	0	0	0	0
9	0	0	0	2	0	0	0	0	0	0	0
10	0	0	0	0	1	0	0	0	0	0	0
11	0	0	0	0	2	0	0	0	0	0	0
12	0	0	0	0	0	1	0	0	0	0	0

13	0	0	0	0	0	2	0	0	0	0	0
14	0	0	0	0	0	0	1	0	0	0	0
15	0	0	0	0	0	0	2	0	0	0	0
16	0	0	0	0	0	0	0	1	0	0	0
17	0	0	0	0	0	0	0	2	0	0	0
18	0	0	0	0	0	0	0	0	1	0	0
19	0	0	0	0	0	0	0	0	2	0	0
20	0	0	0	0	0	0	0	0	0	1	0
21	0	0	0	0	0	0	0	0	0	2	0
22	0	0	0	0	0	0	0	0	0	0	1
23	0	0	0	0	0	0	0	0	0	0	2

Two-factor interaction effects – 220 combinations:

24	1	1	0	0	0	0	0	0	0	0	0
25	1	2	0	0	0	0	0	0	0	0	0
26	2	1	0	0	0	0	0	0	0	0	0
27	2	2	0	0	0	0	0	0	0	0	0
28	1	0	1	0	0	0	0	0	0	0	0
29	1	0	1	0	0	0	0	0	0	0	0
30	2	0	2	0	0	0	0	0	0	0	0
31	2	0	2	0	0	0	0	0	0	0	0
				⋮							
240	0	0	0	0	0	0	0	0	0	1	1
241	0	0	0	0	0	0	0	0	0	1	2
242	0	0	0	0	0	0	0	0	0	2	1
243	0	0	0	0	0	0	0	0	0	2	2

The proportion of zeros to ones to twos is 201:21:21 or roughly a 10:1:1 ratio. In addition to being an inefficient fraction, the solution for effects would involve the inversion of a 243 by 243 matrix. Although the design is saturated and allows estimation of only effects desired, it is inefficient and computationally difficult to analyze.

A saturated plan for mean, main effects, two-factor interactions effect, and three-factor interaction effects would require $1 + 2(11) + 4(11)(10)/2 + 8(11)(10)(9)/2(3) = 1{,}563$ combinations. The ratio of zeros to ones to twos in these combinations would be about 5:1:1. Although this design would be much more efficient, the problem of inverting a 1,563 by 1,563 matrix would pose difficulty.

6. Computer Software for Constructing Orthogonal Fractions of Prime-Numbered Factorials

In this section we introduce a program we have written (named "Wylie") that peforms the calculations described above.

To use Wylie, the investigator first enters the number of factors, levels, and generators (defining contrasts) in the design. He or she then specifies the generators, and the program computes the confounding scheme, resolution, and shortest "words" (i.e., the confounded effects with the fewest factors). At this point, the investigator may modify some of the generators; the confounding scheme, resolution, and shortest words are computed after each change. Finally, when the investigator has decided upon a design, the program lists the design's aliasing structure.

By default, Wylie can handle designs with up to 10 factors at seven levels each, with at most 3000 confounded interactions. These parameters, however, are stored as constants and may easily be modified, depending on design requirements and memory limitations. The program does not handle blocked or folded designs, or designs having factors with unequal numbers of levels.

Wylie was written in THINK Pascal 4.0 on the Apple Macintosh. The code does not rely heavily on Macintosh Toolbox routines or other machine-specific implementations (e.g., there are no graphics), so it is easily portable. Some routines (e.g., the string-manipulation functions and window procedures) are specific to Macintosh THINK Pascal, but most of the code is compatible with standard ANSI Pascal. Source code for Wylie is available from the authors by sending a Macintosh disk.

The following is part of the output for a sample design (the user's responses are italicized):

Enter number of factors: *8*
Enter number of levels: *3*
Enter number of generators: *3*

Enter generator 1: *abcde*
Enter generator 2: *cdefg*
Enter generator 3: *abcdh*

The confounding scheme is
I = abcde
= cdefg = abc2d2e2fg = abf2g2
= abcdh = a2b2c2d2eh = abc2d2efgh = a2b2e2fgh = a2b2cdf2g2h
= eh2 = a2b2efgh2 = cde2fgh2 = c2d2f2g2h2

The design is of resolution 2
The shortest words: eh2

Do you want to replace any generators? (y/n) *y*
Which generator do you wish to replace? (enter a number between 1 and 3) *3*
Enter generator 3: *aefgh*

```
The confounding scheme is
I = abcde
 = cdefg = abc2d2e2fg = abf2g2


= aefgh = a2bcde2fgh = acde2f2g2h = a2bc2d2f2g2h = a2beh
  = bcdf2g2h2 = a2cdh2 = bc2d2eh2 = be2fgh2

The design is of resolution 4
The shortest words: abf2g2  a2beh  a2cdh2

Do you want to replace any generators? (y/n) n

Enter the number of factors (=< 8) in the longest confounded effect you wish to display: 4

The lowest-order resolvable effects are of order t = 2
The aliasing structure for effects of order 1 through 2 are:
 (right-hand side includes only confounded effects of order 2 through 4)

a = a2bf2g2 = beh = cdh2
a2 = bcde = bf2g2 = efgh = abeh = acdh2

b = ab2f2g2 = a2b2eh
b2 = acde = af2g2 = a2eh = c2d2eh2 = e2fgh2

c = a2c2dh2 = bd2eh2
c2 = abde = defg = a2dh2

...

ab = a2b2f2g2 = b2eh = bcdh2
ab2 = a2cde = a2f2g2 = eh = b2cdh2
a2b = b2cde = b2f2g2 = ab2eh
a2b2 = cde = f2g2 = aeh

ac = bceh = c2dh2
ac2 = a2bde = bc2eh = dh2
a2c = bc2de = bcf2g2 = ac2dh2
a2c2 = bde = bc2f2g2 = adh2

...

gh = abf2h = bcdf2 = a2cdg = be2fg2
gh2 = abf2h2 = aefg2 = a2beg
g2h = aefh2 = a2cdg2 = be2f
g2h2 = aef = a2beg2 = be2fh
```

This design took less than one minute of calculation time on a Macintosh SE/30. Smaller designs (e.g., five factors at two levels each, with two generators) can be calculated in a few seconds. A large design with ten factors at five levels each, with four generators, took about five minutes of calculation time.

Work on Wylie has not yet been completed. New features (e.g., the ability to obtain a random ordering of the combinations in the fractional replicate) are being implemented.

References

Anderson, D.A. and W.T. Federer (1975). Possible absolute determinant values for square (0,1)-matrices in fractional replication. *Utilitas Mathematica* **7**, 135-150.

Daniel, C. (1959). Use of half-normal plots in interpreting two-level experiments. *Technometrics* **1**, 311-342.

Federer, W.T. (1955). *Experimental Design — Theory and Application.* Macmillan, New York. Reprinted by Oxford and IBH Publishing Co., New Delhi, 1967.

Kempthorne, O. (1947). A simple approach to confounding and fractional replication in factorial experiments. *Biometrika* **34**, 255-272.

Kempthorne, O. (1952). *The Design and Analysis of Experiments.* John Wiley and Sons, Inc., New York, London, Sydney.

Raktoe, B.L., Hedayat, A. and Federer, W.T. (1981). *Factorial Designs.* John Wiley and Sons, Inc., New York.